



Australian
National
University

ANU College of Engineering and Computer Science
Building 108, North Road
ANU, Canberra 0200, Australia

Final Report for AOARD Grant 124041

What Makes a Message Stick? - The Role of Content and Context in Social Media Epidemics

23/09/2013

Name of Principal Investigator: Lexing Xie

- e-mail address: lexing.xie@anu.edu.au
- Institution : Australian National University
- Mailing Address :
Research School of Computer Science
Building 108, North Road
Canberra, ACT 0200
Australia
- Phone : +61 2 6125 1646
- Fax : +61 2 6125 1646

Period of Performance: 09/29/2012 – 09/28/2013

Overview

When will a message go viral? This is one of the most important questions in analyzing and understanding social media. Our one-year AOARD project tackles this question from two perspectives: understanding individual user preferences, and understanding message popularity from collective user behavior. Our work focuses on building models that predict user behavior and overall popularity. In addition, we also present and analyze observations that explain such behavior from content characteristics and social interactions.

Topic 1: Predicting User Preferences with Fine Grained Social Traits

Inferring the preference of individual users is one important step towards predicting the global content popularity. Content recommendation in social networks poses a complex problem, as users are involved in a rich and complex set of online interactions (e.g., likes, comments and tags for posts, photos and videos) and activities (e.g., favourites, group memberships, interests). While many social collaborative filtering approaches learn from aggregate statistics over this social information, we show that only a small subset of user interactions and activities are actually useful for social recommendation, hence learning which of these are most informative is of critical importance. We design a novel social collaborative filtering approach termed social affinity filtering (SAF) to learn the importance of dozens of interaction types, and thousands of social activities.

On a preference dataset of Facebook users and their interactions with 37,000+ friends collected over a four month period, SAF learns which fine-grained interactions and activities are informative and outperforms state-of-the-art (social) collaborative filtering methods by over 6% in prediction accuracy; SAF also exhibits strong cold-start performance. In addition, we

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 06 NOV 2013		2. REPORT TYPE Final		3. DATES COVERED 29-09-2012 to 28-09-2013	
4. TITLE AND SUBTITLE What Makes a Message Stick? The Role of Content and Context in Social Media Epidemics			5a. CONTRACT NUMBER FA23861214041		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Peter Christen			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Australian National University,Bldg 108 North Road ANU,Canberra ACT 0200,Australia,NA,NA			8. PERFORMING ORGANIZATION REPORT NUMBER N/A		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AOARD, UNIT 45002, APO, AP, 96338-5002			10. SPONSOR/MONITOR'S ACRONYM(S) AOARD		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) AOARD-124041		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT When will a message go viral? This is one of the most important questions in analyzing and understanding social media. Our one-year AOARD project tackles this question from two perspectives: understanding individual user preferences, and understanding message popularity from collective user behavior. Our work focuses on building models that predict user behavior and overall popularity. In addition, we also present and analyze observations that explain such behavior from content characteristics and social interactions.					
15. SUBJECT TERMS sociocultural modeling, sociocultural prediction, Psychology					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 4	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

analyse various aspects of fine-grained social features and show (among many insights) that interactions on video content are more informative than other modalities (e.g., photos), the most informative activity groups tend to have small memberships, and features corresponding to “long-tailed” content (e.g., music and books) can be much more predictive than those with fewer choices (e.g., interests and sports). In summary, this work demonstrates the substantial predictive power of fine-grained social features and the novel method of SAF to leverage them for state-of-the-art social recommendation. For details see attached paper 2 “Social Affinity Filtering: Recommendation through Fine-grained Analysis of User Interactions and Activities”.

Topic 2: Predicting Message Popularity from Content and User Behavior

As a parallel investigation, we directly predict the popularity of a content with two approaches. The first effort looks at video remix on YouTube, and predicts the volume and “longevity” of such community remixes in on-going news events. The second effort looks at the daily viewership of YouTube videos and harvests external signals from Twitter to predict the sudden changes of video views. These two efforts are first of its kind in studying large-scale social remix behavior, and cross-platform influence of popularity. We have collected unique YouTube datasets for evaluating these techniques. They will be made available to the research community after the respective papers are published.

First, we propose visual memes, or frequently re-posted short video segments, for detecting and monitoring latent video interactions at scale. Content sharing networks, such as YouTube, contain traces of both explicit online interactions (such as likes, comments, or subscriptions), as well as latent interactions (such as quoting, or remixing, parts of a video). Visual memes are extracted by scalable detection algorithms that we develop, with high accuracy. We further augment visual memes with text, via a statistical model of latent topics. We model content interactions on YouTube with visual memes, defining several measures of influence and building predictive models for meme popularity. Experiments are carried out with over 2 million video shots from more than 40,000 videos on two prominent news events in 2009: the election in Iran and the swine flu epidemic. In these two events, a high percentage of videos contain remixed content, and it is apparent that traditional news media and citizen journalists have different roles in disseminating remixed content. We perform two quantitative evaluations for annotating visual memes and predicting their popularity. The proposed joint statistical model of visual memes and words out- performs an alternative concurrence model, with an average error of 2% for predicting meme volume and 17% for predicting meme lifespan. For details see attached paper 1 “Tracking Large-Scale Video Remix in Real-World Events”.

Second, we propose a novel method to leverage Twitter features to predict two difficult cases of content popularity on YouTube – the sudden jump in viewcount, and the viewcount of newly uploaded videos. User influence in Twitter and content popularity on YouTube are both very active areas of research, but little attention was devoted to measuring the effects of the former on the latter. We define two classification problems for view-count jump and new video popularity, respectively. We extracted four types of features from Twitter, including information about tweets, twitter user graph, and the interactions that users perform and receive. Prediction performances are reported on thousands of YouTube videos mentioned in a 3-month Twitter feed from 2009. The accuracy for predicting jump improves by 0.10 over a baseline of viewcount history; the accuracy for predicting early popularity

improves by 0.25 over random baseline, where no history is available. These promising results will help a range of applications, including content recommendation on social media, advertising, and others. For an extended summary, see attached abstract 3 "Predicting YouTube Video Viewcount with Twitter Feeds".

List of Publications and Significant Collaborations that resulted from your AOARD supported project.

a) papers published in peer-reviewed journals

[1] "Tracking Large-Scale Video Remix in Real-World Events", Lexing Xie, Apostol Natsev, Xuming He, John Kender, Matthew Hill, John R Smith, IEEE Trans. Multimedia, vol 15 no 6, 1244-1254, Oct 2013. Special Issue on Social Media as Sensors. doi:10.1109/TMM.2013.2264929

b) papers published in peer-reviewed conference proceedings

[2] "Social Affinity Filtering: Recommendation through Fine-grained Analysis of User Interactions and Activities", Suvash Sedhain, Scott Sanner, Lexing Xie, Riley Kidd, Khoi-Nguyen Tran, Peter Christen, ACM Conference on Online Social Networks (COSN 2013), Oct 2013, to appear (full paper, acceptance rate 16%)

c) papers published in non-peer-reviewed journals and conference proceedings
None.

d) conference presentations without papers

- Lexing Xie gave a presentation titled "Understanding Events and Messages Popularity in Media Rich Social Networks" at the Melbourne Social Media workshop, Feb 26, 2013. This workshop is hosted by The Defence Science and Technology Organisation (DSTO) of Australia, and ONR Global (Office of Naval Research, USA) and University of Melbourne.

e) manuscripts to be submitted

[3] "Predicting YouTube Video Viewcount with Twitter Feeds", Honglin Yu, Lexing Xie, Scott Sanner, Abstract enclosed for AOARD report, to be submitted for conference publication in early 2014.

f) provide a list any interactions with industry or with Air Force Research Laboratory scientists or significant collaborations that resulted from this work.

- Lexing Xie visited Microsoft Research Asia (MSRA) in Beijing, August 2013. She presented a seminar titled "Tags, Preferences and Popularity in Social Media", and discussed collaborations in media content analysis, and location-based social networks with MSRA researchers Tao Mei and Xing Xie.
- Lexing Xie and Scott Sanner visited Tsinghua university in Beijing in August and May 2013, respectively. They presented social recommendation and learning knowledge graph in two respective workshops. The workshop is attended by industry partners including

Baidu (major search engine), Tencent (Major microblog and online game portal in China), and Renren (Facebook equivalent in China).

- Lexing Xie attended and presented at AFOSR "Trust and Influence" Program Review in Dayton OH, January 2014. This event was hosted by Joe Lyons, AFOSR PM of the Trust and Influence Program, and attended by multiple AFOSR and Air Force personnel.

Attachments:

Two publications a[1], b[2] and one abstract e[3] to be submitted for publication (waiting for the early 2014 computer science conference cycle) are enclosed.

DD882:

Enclosed.